



**University of  
Zurich**<sup>UZH</sup>

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2019

---

## Decisions against preferences

Messerli, Michael ; Reuter, Kevin

**Abstract:** An agent decides against her preferences, if she considers an option  $x$  better than another option  $y$  but nevertheless decides to do  $y$ . A central tenet of rational choice theory states that individuals do not decide against their preferences, whereby we find two kinds of potential counterexamples in the literature: akrasia, also known as weak-willed decisions, and decisions based on so-called deontic constraints such as obligations or commitments. While there is some empirical evidence that weak-willed choices are a real phenomenon, leading scholars in philosophy of economics debate whether choices based on commitments can be counter-preferential. As far as we know, however, nobody so far has tried to settle this debate empirically. This paper contributes to both debates since we present some empirical evidence that (i) akrasia can also be strong-willed and (ii) choices made on the basis of commitments can indeed be counter-preferential. We will conclude that people can decide against their preferences without being unreasonable

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-174322>

Conference or Workshop Item

Published Version

Originally published at:

Messerli, Michael; Reuter, Kevin (2019). Decisions against preferences. In: The 41st Annual Meeting of the Cognitive Science Society, Montreal, QC, 24 July 2019 - 27 July 2019. CogSci, 2345-2350.

# Decisions Against Preferences

Messerli Michael (m.messerli@sheffield.ac.uk)

Department of Philosophy, Department of Philosophy, 45 Victoria Street, Sheffield, S37QB UK

Reuter Kevin (kevin.reuter@philo.unibe.ch)

Department of Philosophy, Laenggassstrasse 49a  
3012 Bern, Switzerland

## Abstract

An agent decides against her preferences, if she considers an option  $x$  better than another option  $y$  but nevertheless decides to do  $y$ . A central tenet of rational choice theory states that individuals do not decide against their preferences, whereby we find two kinds of potential counterexamples in the literature: *akrasia*, also known as weak-willed decisions, and decisions based on so-called *deontic constraints* such as obligations or commitments. While there is some empirical evidence that weak-willed choices are a real phenomenon, leading scholars in philosophy of economics debate whether choices based on commitments can be counter-preferential. As far as we know, however, nobody so far has tried to settle this debate empirically. This paper contributes to both debates since we present some empirical evidence that (i) *akrasia* can also be strong-willed and (ii) choices made on the basis of commitments can indeed be counter-preferential. We will conclude that people can decide against their preferences without being unreasonable.

**Keywords:** Counter-Preferential Choice; Rational Choice Theory; *Akrasia*; Commitments; Empirical Studies.

## The Putative Irrationality of Deciding against Preferences

A fundamental assumption of most theories of rational choice is that an agent always chooses the option she considers best.<sup>1</sup> Violating this assumption is deemed irrational (Hausman 2012a). This basic tenet applies to both *maximizing* and *optimizing* concepts of rational choice, e.g., if one is a maximizer, one chooses the option that one believes maximizes one's utility. The option that is considered best (or at least not worse than any other) is the one the agent prefers.

What does it mean to *prefer* an option? Savage (1954) takes the notion of preference “in an ordinary mathematical usage by saying that the relation is a simple ordering among acts” (Savage 1954, p. 18). In contrast, philosophers often understand preferences as mental states, e.g., Hausman states “to say that Jill prefers  $x$  to  $y$  is to say that when Jill has thought about everything she takes to bear on how much she values  $x$  and  $y$ , Jill ranks  $x$  above  $y$ ” (Hausman 2012b, p. 34). Of course, Hausman's

notion of preference as *total subjective comparative evaluation* is controversial. Angner (2018), for instance, argues that economists neither *do use* nor *should use* such a conception of preference. However, note that Angner himself accepts that Hausman develops a useful model of preferences. For the purposes of our paper, Hausman's conception is specifically effective because it illustrates why choosing a worse option doesn't seem to make sense. According to such an understanding, an agent follows her preference ordering because she will then do what she values the most, or in other words, what she believes is best for her.

Saying that the agent chooses what she prefers to do is not to say that the agent chooses what is always best for her. In cases of uncertainty, an agent might choose an option that does not maximize her utility since unexpected states of the world might materialize. Agents might also base their preference ranking on false beliefs.<sup>2</sup> They might be mistaken about their preferences, or are simply not able to form a preference ordering (Messerli & Reuter 2017). It seems also false to postulate that agents always take into account all the available information like utilities and probabilities of options (Kahneman & Tversky 2000). Some of these aspects have been used to criticize models of rational choice theory. Nonetheless, these points of criticism do not apply to the fundamental tenet that agents choose what they consider best.

So, are advocates of rational choice theory right that counter-preferential choices do not exist? Or do agents sometimes choose options they consider worse than another, and hence violate this basic assumption of rational choice theory?<sup>3</sup>

<sup>2</sup> Paul (2014), for instance, argues that agents who contemplate so-called *transformative choices* cannot form reasonable beliefs about the content of their experiences, and, thus, cannot make a rational choice. For a critical reply, see, e.g., Reuter & Messerli (2018).

<sup>3</sup> One might object that the assumption that agents choose what they consider best cannot be falsified. Revealed preference theorists, for example, assume that decisions reflect preferences. Consequently, there is no conceptual gap between a person's preferences and the actions she decides to perform. However, on most philosophers' interpretation, expected utility represents the strength of an agent's preference for the outcome, where preferences are understood as psychological states. Given this inter-

<sup>1</sup> Some advocates of satisficing concepts of rational choice would disagree. If one is a satisficer, one settles for any alternative one considers satisfactory (Simon 1953; Slote 2004).

As far as we know, there are at least two kinds of potential counterexamples challenging this assumption. First, decisions can be *deontically constrained*, a technical term used to refer to constraints that arise when morality requires us to act in ways that are contrary to self-interest (see, e.g., Heath 2008). Put differently, a deontic constraint can be understood as a form of duty or rule that makes people refrain from the pursuit of individual advantage. That may sound fairly abstract, but we all are familiar with situations in which we act because we have given a promise, not because we actually prefer to act that way. We will come back to this issue in the General Discussion when discussing whether commitments can be counter-preferential.

Second, weak-willed decisions (or akratic decisions) are also potential counterexamples violating the assumption that agents choose what they consider best. To illustrate akratic decisions, take the case of Lewd Larry: Larry believes that staying in his room and staying faithful to his girlfriend is better than having an affair with his flatmate Jackie, but then finds himself trying to seduce her. Davidson (1970) defines weak-willed actions as follows:<sup>4</sup> In doing *y* an agent acts weak-willed if and only if: (i) the agent does *y* intentionally; (ii) the agent believes there is an alternative action *x* open to him; (iii) the agent judges that, all-things considered, it would be better to do *x* than to do *y*. Lewd Larry seems to fulfill all the requirements for being weak-willed. Weak-willed decisions not only seem to violate a fundamental assumption of rational choice theory, the intuitive plausibility of Lewd Larry demonstrates that weak-willed actions are real. However, most scholars at least agree with advocates of rational choice theory that such a decision is irrational: he should not have decided to seduce Jackie, given his belief it is not his best option.

In the rest of this paper, we do three things: First, we describe a case illustrating a violation of the aforementioned fundamental assumption of rational choice theory, which is not (at least not intuitively) an akratic decision, and, we present a first experiment showing that these cases are real. Second, we discuss an objection against our study and results, and we counter this objection using a second study. Third, taking ideas from Amartya Sen, we argue that we have good reasons to believe that such decisions are reasonable choices. In other words, such choices can be understood as acting out of commitment, whereby the commitment is counter-preferential.

pretation of rational choice theory, the assumption that agents actually choose options they consider worse than another is empirically testable.

<sup>4</sup> For discussions on the concept of akrasia as well as criticisms on Davidson's definition, see, e.g., Mele (1991), Holton (1999), and May & Holton (2012).

## Experimental Study 1

Examples in which an agent decides against her preferences almost always seem to have the following pattern. The agent values *x* more than *y*, and hence prefers *x* to *y*, but "lower" desires triggered by lust or sloth, move the agent to do *y*. It need not be the case, however, that an agent decides against her preferences only if she is weak-willed. An agent might value *x* more than *y*, and hence prefers *x* to *y*, but is moved by his "higher" commitments or obligations to do *y*. To illustrate such a case, take the following example. Today, a colleague of yours has asked you whether you would help him move some furniture, and you agreed to be at his place the next morning. The next morning, however, friends of yours ask you whether you would like to join them for a beautiful day at the lake. It seems at least possible that in such a situation, you value going to the lake more than helping your colleague move furniture. Nonetheless, you decide to be at your colleague's place and help him move furniture. Note that similar to typical weak-willed decisions, you might loathe the fact that you have acted contrary to what you considered the best option. While such actions satisfy Davidson's definition of being weak-willed, it seems highly odd to call them weak-willed.<sup>5</sup>

The decision we described above violates the basic tenet of rational choice theory just as much as weak-willed decisions. The agent does not maximize her utility by choosing an option she considers worse than an available alternative. But are these decisions actually real? Or are they mere figments of philosophers' imaginations? The following experiment strongly suggests that these decisions are part of many people's reality.

## Methods

120 participants were recruited on Amazon Mechanical Turk and paid a small fee for their participation. 5 participants were excluded for not having completed the survey. The remaining 115 participants (51 women,  $M_{age} = 39.09$ ,  $SD = 15.69$ ) all indicated that they were native English speakers. All participants were randomly assigned to one of three conditions, two test conditions (*Acquaintance*, *Colleague*) and one control condition. The

<sup>5</sup> Davidson discusses similar, so-called incontinent cases, in which a person follows a duty or principle when doing *y* intentionally (e.g., getting up and brushing her teeth), although all-things considered, she judges *x* to be better than *y*, e.g., it is more pleasurable to stay in bed. However, there is an important difference between our cases and Davidson's incontinent cases. In Davidson's examples, the agent does not believe she has very good reasons to follow a certain duty or principle, e.g., the agent reasons that her teeth are very strong anyway. In our examples, the agent is likely to believe that she has good reasons to keep a promise. In other words, the agent believes that helping a colleague move is the right thing to do. This contrast explains why Davidson believes that in incontinent cases an agent cannot understand herself and that she recognizes something absurd in her intentional behavior, while this would not be true in our case.

vignettes of the two test conditions read as follows:

**Test condition** Imagine that an acquaintance (a colleague) of yours asks you whether you would help him move some furniture and household appliances into his new apartment. You agree to be at his place at 10am the next day. The next morning, it is a beautiful warm summer day. At 8am you get a call from friends who ask you whether you would like to join them for a nice day at a lake. All things considered and independent of how you decide in the end, how do you value each of the two options:

- Spending the day at the lake and tell my acquaintance (colleague) that I cannot come.
- Moving furniture and household appliances and tell my friends that I cannot join them.<sup>6</sup>

The vignette for the control condition read:

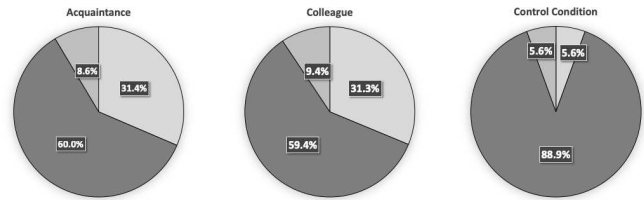
**Control condition** Imagine that you plan your yearly holidays. On the one hand, you could go to the seaside and spend a week relaxing at the beach. On the other hand, you could book a trip to a city you have not seen before and experience some cultural highlights. All things considered and independent of how you decide in the end, how do you value each of the two options:

- Spending the holidays on the beach and not going to a city.
- Spending the holidays in a city and not going to the beach.

After the participants rated both options, they were then directed to the second question reading:

**Decision Question** You have just valued each of the two options. But how do you decide in the end? Please tell us what you will do:

For the two test conditions, the participants were presented with two options: (1) I choose to go to the lake and tell my acquaintance (colleague) I cannot come. (2) I choose to move furniture and household appliances and



**Figure 1** Responses in % to the two test conditions and the control condition. Dark grey depicts the percentage of participants who would decide in line with their preferences. Light grey represents the percentage of participants who would decide against their preferences. A few participants (medium grey) indicated an equal preference for both options.

tell my friends I cannot join them. In the control condition, the options were: (1) I choose to go to the seaside and spend a week relaxing at the beach. (2) I choose to go to a city and experience some cultural highlights. Participants had to choose which decision they would take.

## Results

The results of people's responses are summarized in Figure 1 above. In both the *Acquaintance* as well as the *Colleague* condition, around 31% of the participants who decided in favor of helping to move furniture, considered going to the lake more valuable. The response profiles were significantly different between the test condition *Acquaintance* and control,  $\chi^2 = 8.70, p = 0.013$ , as well as *Colleague* and control,  $\chi^2 = 8.64, p = 0.013$ .

## Discussion

The results indicate that a substantial portion of the participants would decide in favor of a less valuable option when they consider the scenario we presented them with. Simply put, the results suggest that there are situations in which many people will decide against their own preferences. The study was not designed to investigate which percentage of people are likely to make a decision against their preferences. Obviously, the scenarios were quite specific and for many the situation did not even present them with a "difficult" choice. Thus, it is likely that for many more people than just the recorded 31%, there exist choices in which option  $x$  is preferred but they still decide in favor of option  $y$ . Of course, for most decisions, people's choices will nicely align with their preferences. In fact, the control condition was specifically designed as a base rate for decisions in which preferences are the sole determiner of the decision in question. The significant differences between the test condition and the control demonstrates, however, that not all decisions are like that. Other factors may determine which option we are going to choose.

Before we discuss a possible explanation for the recorded data, let us first address the most obvious objection against our study. In order to counter this objection, we then briefly present the results of a second study.

<sup>6</sup> Both options were presented in randomized order and participants were asked to rate the value of each option on an 11-point Likert scale anchored at 0 meaning "Not at all valuable" and 10 meaning "Extremely valuable". Which concept of utility is relevant here? Importantly, we do not understand utility as a more precise ranking than an ordinal one (e.g. *cardinal measure* or *ratio scale*). The strength of a value judgement can be understood in purely ordinal terms, respectively, the experiment is perfectly consistent with an ordinal interpretation. If a participant evaluates two options within this scale, e.g.  $a = 9$  and  $b = 3$ , this simply means that he or she ranks  $a$  above  $b$ . In other words, the only information these numbers provide is that an agent prefers  $a$  to  $b$  without saying how much he or she values  $a$  more than  $b$ .

## Objection

The experiment reveals a potential decision against one's preferences, only if people gave all-things-considered value ratings when considering their options. It is indeed possible that some people merely considered the positive value of spending a day on the lake without considering the negative value of telling one's acquaintance or colleague that one is not available for moving after all. If that were the case, then it would not surprise to see decisions made against one's *rated* preferences.

We do not believe, however, that this is a likely possibility. When we asked the participants to rate the value of the options, we specifically named the positive as well as the negative aspect of the choice, e.g., one of the options read: "Spending the day at the lake and tell my colleague that I cannot come." Moreover, the number of participants who would decide against their own value judgements might even be greater, because some participants might have self-censored themselves so as to appear to be consistent when making a decision.

However, one might insist on the ambiguity of the term "value", respectively, that we and the participants do not refer to the same concept here. It is our understanding that the concept of value can be understood in terms of the agent's ends and desires. "I judge that *a* is more valuable than *b*" means that I believe that *a* is more valuable than *b* in terms of my ends and desires.<sup>7</sup> Now the objection that arises is that participants must have some different concept of value in mind, because there is not only the end of enjoying a great day at the lake but also the end of helping other people, which is obviously more important to them. If the objection stands, participants do in fact decide in line with their values and do not decide counter-preferentially.<sup>8</sup>

The objection we raised should be taken seriously. We

<sup>7</sup> In accordance with rational choice theory, we do not make any proposal concerning the content of these ends. This means that we recognize no distinction between goals such as making a million dollar, helping other people and being a sadist. Also note that there are no implications regarding risk-taking. The value judgement that *a* is more valuable than *b* might be risk-neutral such as in standard approaches or risk-averse such as in *prospect theory*.

<sup>8</sup> One way of testing the objection would be to further specify the alternatives, e.g., instead of stating one of the options as "Spending the day at the lake and tell my colleague that I cannot come," we could state "Spending the day at the lake and break my promise to my colleague". The reason why we opted for a different way to tackle the objection is that "breaking a promise" or "breaking a commitment" (we will come back to the role of commitments in the General Discussion) is a very negative trigger. The wording "tell my colleague that I cannot come" is relatively neutral in this regard. However, we agree that the empirical evidence for decisions against preferences would be even greater if the negative aspects of a decision would be highlighted even further. In a follow-up study, we plan not only to investigate a larger variety of experimental stimuli but also the impact of the exact wording on the empirical effect.

have, therefore, conducted a second experiment where we first explained to participants which concept of value is involved. We will see that our results are robust, even if we change the experimental setting in this way.

## Experimental Study 2

In order to address the objection stated above, we decided to rerun both test conditions (*Acquaintance*, *Colleague*) to see whether the results would change or remain robust. If the objection is correct, then we should see a substantial drop in the percentages of people who indicate decisions that go against their preferences.

## Methods

100 participants were recruited on Amazon Mechanical Turk and paid a small fee for their participation. 2 participants were excluded for not having completed the survey. The remaining 98 participants (48 women,  $M_{age} = 36.92$ ,  $SD = 12.38$ ) all indicated that they were native English speakers. All participants were randomly assigned to one of two conditions (*Acquaintance*, *Colleague*). The vignettes of the two conditions were exactly the same as the vignettes of the test conditions in Experiment 1 with one exception: after participants had given their consent to this study, they were informed about the task ahead in the following manner.

**Instructions** On the next screen we will ask you to value certain events. Before you do so, please consider the following example: Imagine you have to value a one week trip to Europe. On the positive side there might be aspects like relaxing, eating new and exciting food, being able to tell your friends of an amazing trip when you are back, etc. On the negative side there might be aspects like being jetlagged, longing for your loved ones at home, missing an important meeting at work, etc. Thus, if you value an option or an event, you take into account all its positive and negative aspects and then make an overall judgement.

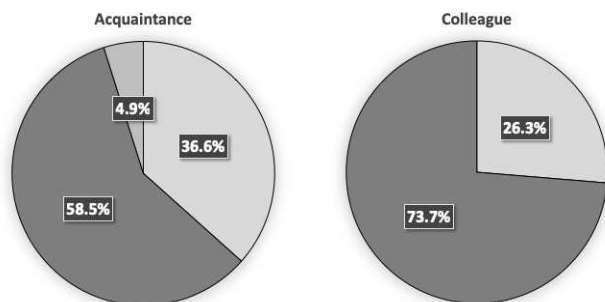
After these instructions, participants rated both options (see Experiment 1 above), and then answered the decision question (see also Experiment 1 above).

## Results

In the *Acquaintance* condition, 36.6% of the participants who decided in favor of helping to move furniture considered going to the lake more valuable. In the *Colleague* condition, 26.3% of the participants who decided in favor of helping to move furniture considered going to the lake more valuable. The results of people's responses are summarized in Figure 2 below.

## Discussion

The data we received in Experiment 2 are highly similar to those we collected in Experiment 1. While the percentage of people who decided against their preference in the



**Figure 2** Responses in % to the two conditions. Dark grey depicts the percentage of participants who would decide in line with their preferences. Light grey represents the percentage of participants who would decide against their preferences. A few participants (medium grey) indicated an equal preference for both options.

*Acquaintance* condition rose from 31.4% to 36.6%, the percentages in the *Colleague* condition decreased from 31.3% to 26.3%. Thus, overall the results of Experiment 1 were very robust. It seems therefore very likely, that the participants in Experiment 1 entertained a notion of *value* not only similar to the one in Experiment 2, but also of the right kind.

## General Discussion

The studies suggest that people often make decisions in favor of options they consider less valuable.<sup>9</sup> If our results are correct, there are crucial implications for both the discussion on *akrasia* and the debate on *rational choice*, respectively, the connection between rational choice and *deontic constraints*.

First, let us briefly mention the implication for *akrasia*. Importantly, some philosophers mention that *akrasia* can also be strong-willed (e.g., see Holton 1999; Yao 2017). However, as far as we know, (i) nobody thus far has interpreted such cases as actions out of commitment, (ii) there is currently no empirical evidence for such cases, and, (iii) other cases of strong-willed *akrasia* are related to violations of resolutions and preference change. In contrast, our example illustrates a case in which no such additional machinery is necessary.

Second, our results are crucial for rational choice and the debate on commitments. Most rational choice theorists are likely to consider counter-preferential decisions as unreasonable or irrational, similar to typical weak-willed decisions. However, while people in weak-willed decisions usually act out of lower desires, we have des-

cribed a case in which people seem to be rather strong-willed when acting against their most valued option. Does this difference allow us to frame such decisions as reasonable? This largely depends on what ultimately motivates people to decide against preferences in *strong-willed* decisions. A possible explanation of such decisions takes into account the importance of commitments. After all, many people are likely to decide to help their acquaintance or colleague move furniture because they have committed themselves to do so, not because they like moving furniture. However, shouldn't these commitments be reflected in peoples' evaluations of the two alternatives? According to Sen (1977) this need not be the case.

Sen distinguishes three kinds of motivations: narrow self-interest, sympathy, and commitment. Both, narrow self-interest as well as sympathy, directly affect a person's own welfare and should be reflected in people's value judgements. In contrast, Sen (1977, p. 326) characterizes *commitments* as altruistic attitudes towards others. Accordingly, a person who acts out of commitment chooses an option that she considers the right thing to do, even if that option is less preferable than an alternative. Sen admits that within the framework of rational choice theory, there is no place for a notion like commitment because it does not lead to any difference in terms of one's expected advantage.<sup>10</sup> Speaking purely in terms of rational choice theory, decisions against preferences, are therefore irrational. That said, Sen's theoretical work on commitments has caused a lot of attention, because it seems that people who act out of commitment, are *irrational* only in the skewed notion of rational choice theory. At least, intuitively, it seems that people who act against their preferences but in favor of an option they consider the right thing to do, are reasonable agents.

Given the importance of Sen's contribution, some philosophers have started to question Sen's depiction of commitments as factors that may have a motivating force beyond expected advantage. Contra Sen, Hausman (2007) argues that we need to distinguish among the variety of factors responsible for agents' preferences, rather than distinguish between preferences and commitments. According to his view, commitments are not counter-preferential but rather influence all-things considered judgements. Thus, while Sen believes commitments can directly determine our choices, Hausman argues that they do so only via preferences. As far as we know, the role of commitments in the decision making process has not yet been empirically investigated.<sup>11</sup> And

<sup>9</sup> One might object that there is a gap between the participants *rated* preferences and their real decisions or real behavior. Put differently, participants are not actually making a decision but provide inconsequential responses after reading abstract descriptions of some options. However, while some studies have shown an inconsistency between people's rated preferences and real behavior, a variety of empirical studies have also shown high consistency between people's ratings and their behavior. Importantly, we are not aware of any empirical or theoretical arguments why people systematically deviate in our respective context.

<sup>10</sup> It is important to keep in mind here that Sen distinguishes different notions of preferences. The two most important ones are (i) preference as (revealed) choice ranking and (ii) preference as expected advantage ranking.

<sup>11</sup> Note that we do not claim that there's no empirical research on commitments. See, e.g., Székely & Michael

therefore, we do not yet know whether Sen or Hausman are right. However, our experiment may provide a first step to settle this debate. According to our results, it seems that commitments may sometimes influence our choices directly and not via preferences. At a minimum, opponents of Sen would need to explain why some people decide to move furniture, even if the other option is considered better, all-things-considered.

Before we conclude, let us briefly mention one other account that could be drawn upon to explain our data. Heath (2008) argues that theories of rational choice can be modified in order to incorporate rule following behavior. His model distinguishes between one's desire for an outcome (its expected utility) and how appropriate the outcome is (the normative appropriateness of that outcome). The basic idea is that an agent's utility function combines two things: Getting the best outcome and doing the right thing. According to Heath's approach, participants would distinguish two stages. First, they would rank permissible actions as more or less appropriate. Second, they would add these values to the expected utilities. It would take more experiments to find out whether participants indeed proceed in the way suggested by Heath. In any case, Sen's account provides a straightforward explanation of our data.

## Conclusion

In closing, let us summarize what we have done. Adherents of rational choice theory assume that agents choose the option they consider best. In this paper, we have discussed a case that violates this basic assumption. Crucially, we have not merely relied on our own intuitions of whether such a case is real, but conducted two studies, the results of which strongly suggest that many people make decisions against their preferences. Some might argue that this case is just one out of many showing rational choice theory to be mistaken. In particular, weak-willed decisions have been largely accepted to be real-world cases in which agents act contrary to their best judgements. However, weak-willed decisions can be distinguished from our case study in two important respects: First, while in weak-willed decisions, people act out of their lower desires, our case shows that people can decide against their preferences by being strong-willed. Second, at least according to Sen's account, there are good reasons to believe, agents may act against their preferences but at the same time make a reasonable choice. Our results provide evidence that Sen is right, respectively, that commitments can be counter-preferential.

## Acknowledgments

We would like to thank Catherine Herfeld for her very helpful comments on this paper. Michael Messer-

li was supported by the SNF Postdoc Mobility Grant P2SKP1\_171776. Kevin Reuter was supported by the SNF project Affective Mind, Grant No. 169484.

## References

- Angner, Erik (2018). What Preferences Really Are, in: *Philosophy of Science* 85 (4), 660-681.
- Davidson, D. (1970). How Is Weakness of the Will Possible?, in *Essays on Actions and Events*, Oxford: Clarendon Press.
- Hausman, D. (2007). Sympathy, Commitment, and Preference, in: *Rationality and Commitment*, F. Peter and B. Schmid (ed.), Oxford: Oxford University Press.
- Hausman, D. (2012a). *Philosophy of Economics*, The Stanford Encyclopedia of Philosophy (Winter 2013 Edition), Edward N. Zalta (ed.).
- Hausman, D. (2012b). *Preference, Value, Choice, and Welfare*, Cambridge: Cambridge University Press.
- Heath, J. (2008). *Following the Rules. Practical Reasoning and Deontic Constraint*, Oxford: Oxford University Press.
- Holton, R. (1999). Intention and Weakness of Will, in: *Journal of Philosophy* 96 (5), 241-262.
- Kahneman, D. & Tversky, A. eds. (2000). *Choices, Values and Frames*, New York: Cambridge University Press and the Russell Sage Foundation.
- May, J. & Holton, R. (2012). What in the World is Weakness of Will?, in *Philosophical Studies* 157 (3), 341-360.
- McClennen, E. (1990). *Rationality and Dynamic Choice*, Cambridge: Cambridge University Press.
- Mele, A. (1991). Akratic Action and the Practical Role of Better Judgement, in: *Pacific Philosophical Quarterly*, 72 (1), 33-47.
- Messerli, M. & Reuter, K. (2017). Hard Cases of Comparison, in: *Philosophical Studies* 174 (9), 2227-2250.
- Paul, L.A. (2014). *Transformative Experience*. Oxford: Oxford University Press.
- Reuter, K. & Messerli, M. (2018). Transformative Decisions, in: *Journal of Philosophy*, 115 (6), 313-335.
- Simon, H. (1953). *A Behavioural Model of Rational Choice*. Santa Monica: Rand.
- Sen, A. (1977). Rational Fools: A Critique of the Behavioral Foundations of Economic Theory, in: *Philosophy & Public Affairs* 6 (4), 317-344.
- Slote, M. (2004). Two Views of Satisficing in: *Satisficing and Maximizing*, Byron, M. (ed.), Cambridge: Cambridge University Press.
- Székely, M. & Michael, J. (2018). Investing in commitment: Persistence in a joint action is enhanced by the perception of a partner's effort, in: *Cognition*, 174, 37-42.
- Yao, V. (2017). Strong-willed Akrasia. *Oxford Studies in Agency and Responsibility*, 4, 6-27.

---

(2018). As far as we know, there has been no empirical research on the question of whether commitments can be counter-preferential.